

NO-A191 919

NAVAL C3 DISTRIBUTED TACTICAL DECISIONMAKING(U)
MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR
INFORMATION AND DECISION SYSTEMS M ATHANS ET AL

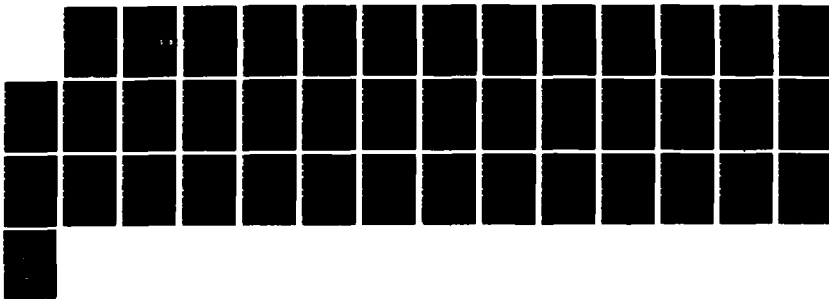
1/1

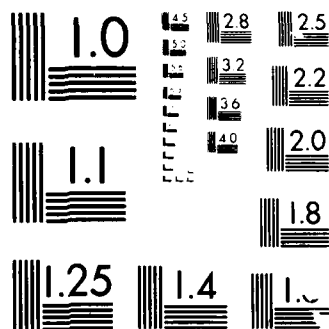
UNCLASSIFIED

22 JAN 88 LIDS-IR-1741 N00014-84-K-0519

F/G 25/5

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A191 919

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. REPORT'S CATALOG NUMBER
4. TITLE (and Subtitle) NAVAL C ³ DISTRIBUTED TACTICAL DECISIONMAKING		5. TYPE OF REPORT & PERIOD COVERED 1 Oct. 1987-31 Dec. 1987
		6. PERFORMING ORG. REPORT NUMBER LIDS-IR-1741
7. AUTHOR(s) Michael Athans Alexander H. Levis		8. CONTRACT OR GRANT NUMBER(s) N00014-84-K-0519
9. PERFORMING ORGANIZATION NAME AND ADDRESS Laboratory for Information and Decision Systems Massachusetts Institute of Technology Cambridge, MA. 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 649-003
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217-5000		12. REPORT DATE Jan. 22, 1988
		13. NUMBER OF PAGES 1+25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Dr. Neil Gerr Mathematical Sciences Division Office of Naval Research Arlington, VA. 22217-5000		15. SECURITY CLASS. of this report UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT of this Report Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT of the abstract entered in Block 20, if different from Report DTIC ELECTE FEB 18 1988 S D		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Command and Control, Distributed Decisionmaking, Organization Theory, Petri Nets.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Progress on eight research problems addressing distributed tactical decisionmaking is described.		

DD FORM 1473
JAN 73

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LABORATORY FOR INFORMATION AND DECISION SYSTEMS**

Cambridge, Massachusetts 02139

OSP NUMBER 95442

LIDS-IR-1741

QUARTERLY TECHNICAL REPORT

for the period

1 October 1987 to 31 December 1987

for

NAVAL C³ DISTRIBUTED TACTICAL DECISION MAKING

**Contract Number N00014-84-K-0519
Work Unit NR 649-003**

Submitted to:

**Dr. N. Gerr, Code 1111SP (3 copies)
Dr. H. L. Hawkins, Code 1111SP
Dr. R. Madan, Code 414
Dr. W. R. Vaughan, Jr.**

**Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217-5000**

Submitted by:

**Michael Athans
Alexander H. Levis**

January 22, 1988

88 2 17 004

NAVAL C³ DISTRIBUTED TACTICAL DECISIONMAKING

1. PROJECT OBJECTIVES

The objective of the research is to address analytical and computational issues that arise in the modeling, analysis and design of distributed tactical decisionmaking. The research plan has been organized into two highly interrelated research areas:

- (a) Distributed Tactical Decision Processes;
- (b) Distributed Organization Design.

The focus of the first area is the development of methodologies, models, theories and algorithms directed toward the derivation of superior tactical decision, coordination, and communication strategies of distributed agents in fixed organizational structures. The framework for this research is normative.

The focus of the second area is the development of a quantitative methodology for the evaluation and comparison of alternative organizational structures or architectures. The organizations considered consist of human decisionmakers with bounded rationality who are supported by C³ systems. The organizations function in a hostile environment where the tempo of operations is fast; consequently, the organizations must be able to respond to events in a timely manner. The framework for this research is descriptive.

2. STATEMENT OF WORK

The research program has been organized into seven technical tasks -- four that address primarily the theme of distributed tactical decision processes and three that address the design of distributed organizations. An eighth task addresses the integration of the results. They are:

- 2.1 Real Time Situation Assessment: Static hypothesis testing, the effect of human constraints and the impact of asynchronous processing on situation assessment tasks will be explored.
- 2.2 Real Time Resource Allocation: Specific research topics include the use of algebraic structures for distributed decision problems, aggregate solution techniques and coordination.



Availability Codes	
Dist	Avail and/or Special
A-1	

- 2.3 Impact of Informational Discrepancy: The effect on distributed decisionmaking of different tactical information being available to different decisionmakers will be explored. The development of an agent model, the modeling of disagreement, and the formulation of coordination strategies to minimize disagreement are specific research issues within this task.
- 2.4 Constrained Distributed Problem Solving: The agent model will be extended to reflect human decisionmaking limitations such as specialization, limited decision authority, and limited local computational resources. Goal decomposition models will be introduced to derive local agent optimization criteria. This research will be focused on the formulation of optimization problems and their solution.
- 2.5 Evaluation of Alternative Organizational Architectures: This task will address analytical and computational issues that arise in the construction of the generalized performance-workload locus. This locus is used to describe the performance characteristics of a decisionmaking organization and the workload of individual decisionmakers.
- 2.6 Asynchronous Protocols: The use of asynchronous protocols in improving the timeliness of the organization's response is the main objective of this task. The tradeoff between timeliness and other performance measures will be investigated.
- 2.7 Information Support Structures: In this task, the effect of the C³ system on organizational performance and on the decisionmaker's workload will be studied.
- 2.8 Integration of Results: A final, eighth task, is included in which the various analytical and computational results will be interpreted in the context of organizational bounded rationality.

3. STATUS REPORT

In the context of the first seven tasks outlined in Section 2, a number of specific research problems have been formulated and are being addressed by graduate research assistants under the supervision of project faculty and staff. Research problems which were completed prior to or were not active during this last quarter have not been included in the report.

3.1 DISTRIBUTED TEAM HYPOTHESIS TESTING WITH EXPENSIVE COMMUNICATIONS

Background: In Command-Control-and-Communication (C^3) systems multiple hypothesis testing problems abound in the surveillance area. Targets must be detected and their attributes must be established; this involves target discrimination and identification. Some target attributes, such as location, are best observed by sensors such as radar. More uncertain target locations are obtained by passive sensors, such as sonar or IR sensors. However, target identity information requires other types of sensors (such as ESM receivers, IR signature analysis, human intelligence etc). As a consequence in order to accurately locate and identify a specific target out of a possibly large potential population (including false targets) one must design a detection and discrimination system which involves the fusing of information from several different sensors generating possibly specialized information about the target. These sensors may be collocated on a platform (say a ship in a Naval battle group) or be physically dispersed as well (ESM receivers exist in every ship, aircraft, and submarine). The communication of information among this diverse sensor family may be difficult (because of EMCON restrictions) and is vulnerable to enemy countermeasure actions (physical destruction and jamming). It is this class of problems that motivates our research agenda.

To put it another way the fusion of information derived from dispersed sensors and decision nodes requires communication. To discourage nonessential communication we would like to put a price on each transmitted bit. In this manner, extensive communications would occur only if the decision warrants them.

Research Goals: We are conducting research on distributed multiple hypothesis testing using several decision-makers, and teams of decision-makers, with distinct private information and limited communications. This is the simplest possible non-trivial distributed decision problem, whose centralized counterpart is well understood and straight-forward to compute. The goal of this research is to *unify* our previous research in situation assessment, distributed hypothesis testing, and impact of informational discrepancy; and to *extend* the methodology, mathematical theory and computational algorithms so that we can synthesize and study more complex organizational structures. The solution of this class of basic research problems will have impact in structuring the distributed architectures necessary for the detection, discrimination, identification and classification of attributes of several targets (or events) by a collection of distinct sensors (or

dispersed human observers).

The objective of the distributed organization will be the resolution of several possible hypotheses based on many uncertain measurements. Each hypothesis will be characterized by several attributes. Each attribute will have a different *degree of observability* to different decision makers or teams of decision makers; in this manner, we shall model different specialization expertise associated with the detection and resolution of different phenomena. Since each hypothesis will have several attributes, it follows that in order to reliably confirm or reject a particular hypothesis, two or more decision-makers (or two or more *teams* of decision-makers) will have to pool and fuse their knowledge.

Extensive and unnecessary communication among the decision-makers will be discouraged by explicitly assigning costs to certain types of communication. In this manner, we shall seek to understand and isolate which communications are truly vital in the organizational performance; the very problem formulation will discourage communications whose impact upon performance is minimal. Quantitative tradeoffs will be sought.

We stress that we shall strive to design distributed organizational architectures in which teams of teams of decision-makers interact. For example, a team may consist of a *primary* decision-maker together with a *consulting* decision-maker -- the paradigm used by Papastavrou and Athans.

The methodology that we plan to employ will be mathematical in nature. To the extent possible we shall formulate the problems as mathematical optimization problems. Thus, we seek *normative solution concepts*. To the extent that human *bounded rationality* constraints are available, these will be incorporated in the mathematical problem formulation. In this case, the nature of the results will correspond to what is commonly referred to as *normative/descriptive* solutions. Therefore, we visualize a dual benefit of our basic research results. From a purely mathematical point of view, the research will yield nontrivial advances to the distributed hypothesis-testing problem; an very difficult problem from a mathematical point of view. From a psychological perspective, we hope that the normative results will suggest *counterintuitive* behavioral patterns of -- even perfectly rational -- decision-makers operating in a distributed tactical decision-making environment; these will set the stage for designing empirical studies and experiments and point to key variables that should be observed, recorded and analyzed by cognitive scientists. From a military C^3 viewpoint, the results will be useful in structuring distributed architectures for the

surveillance/discrimination function.

Progress during the past quarter: In the past quarter we completed the investigation of the problem of *ternary* hypothesis testing by a team of two cooperating decision makers; communication between the two decision-makers is costly and consists of a finite alphabet. The problem is to distinguish among three different hypotheses. Each decision-maker obtains an uncertain measurement of the true hypothesis. The so-called *primary* decision-maker has the option of making the final team decision or consulting, at a cost, the *consulting* decision-maker. The consulting decision-maker is constrained to provide information using a ternary alphabet. The team objective is to minimize the probability of error together with the communications cost (if any). Mr. Papastavrou, under the supervision of Prof. Athans, has derived all necessary equations. However, due to the severe complexity of these equations, we decided not to write the necessary software for their solution at the present time.

Mr. Papastavrou and Professor Athans have initiated the investigation of a class of distributed decision problems originally analyzed by L. Ekchian in his Ph.D. thesis (1983). Consider the problem of binary hypothesis testing by two decision makers (DMs) connected in tandem. The "upstream" DM communicates his conclusion to the "downstream" DM who then blends his measurement with the "upstream" decision, and generates the final decision for the team. The quality of each DM can be quantified by his receiver operating characteristic (ROC) curve. Dominance of the ROC curves can be used to indicate that a particular DM is clearly better than the other one. Ekchian had posed the (reasonable) conjecture that the better DM should be the downstream one. We have been able to verify this conjecture for a class of gaussian problems, and we are attempting to either prove the conjecture in general, or construct a counter-example. This line of inquiry is important because it would point out how relative expertise of DMs should impact organizational design.

Mr. Pothiwala and Professor Athans have also examined the above problem under the assumption that the upstream DM is allowed to communicate with more than two bits his tentative decision to the downstream DM. We seek to understand the value of each additional bit of communicated information to the overall improvement of the distributed team objective (e.g. the weighted probability of error).

Documentation: We have started a paper on the binary hypothesis testing problem for presentation at the upcoming JDL C2 Symposium in June 1988. An abstract has been submitted.

3.2 DISTRIBUTED HYPOTHESIS TESTING WITH MANY AGENTS

Background: The goal of this research project is to develop a better understanding of the nature of the optimal messages to be transmitted to a central command station (or fusion center) by a set of agents (or sensors) who receive different information on their environment. In particular, we are interested in solutions of this problem which are tractable from the computational point of view. Progress in this direction has been made by studying the case of a large number of agents. Normative/prescriptive solutions are sought.

Problem Statement: Let H_0 and H_1 be two alternative hypotheses on the state of the environment and let there be N agents (e.g. intelligent sensors) who possess some stochastic information related to the state of the environment. In particular, we assume that each agent i observes a random variable y_i with known conditional distribution $P(y_i|H_j)$, $j = 0, 1$, given either hypothesis. We assume that all agents have information of the same quality, that is, the random variables are identically distributed. Each agent transmits a binary message to a central fusion center, based on his information y_i . The fusion center then takes into account all messages it has received to declare hypothesis H_0 or H_1 true. The problem consists of determining the optimal strategies of the agents as far as their choice of message is concerned. This problem has been long recognized as a prototype problem in team decision theory: It is simple enough so that analysis may be feasible, but also rich enough to allow nontrivial insights into optimal team decision making under uncertainty.

Results: This problem has been studied by Prof. J. Tsitsiklis. Past results [1-2] can be summarized as follows: Under the assumption that the random variables y_i are conditionally independent (given either hypothesis), it is known that each agent should choose his message based on a likelihood ratio test. Nevertheless, we have constructed examples which show that even though there is a perfect symmetry in the problem, it is optimal to have different agents use different thresholds in their likelihood ratio tests. This is an unfortunate situation, because it severely complicates the numerical solution of the problem (that is, the explicit computation of the decision threshold of each agent). Still, we have shown that in the limit, as the number of agents becomes large, it is asymptotically optimal to have each agent use the same threshold. Furthermore, there is a simple effective computational procedure for evaluating this single optimal

threshold.

We have also shown that if each agent is to transmit K-valued, as opposed to binary messages, then still each agent should use the same decision rule, when the number of agents is large. Unfortunately, however, the computation of this particular decision rule becomes increasingly harder as K increases.

We have also investigated the case of M-ary ($M > 2$) hypothesis testing and constructed examples showing that it is better to have different agents use different decision rules, even in the limit as $N \rightarrow \infty$. Nevertheless, we have shown that the optimal set of decision rules is not completely arbitrary. In particular, it is optimal to partition the set of agents into at most $M(M-1)/2$ groups and, for each group, each agent should use the same decision rule. The decision rule corresponding to each group and the proportion of the agents assigned to each group may be determined by solving a linear programming problem, at least in the case where the set of possible observations by each agent is finite.

Finally, results have been obtained which cover the Neyman-Pearson (as opposed to Bayesian) version of the problem, in the case of $M=2$ hypothesis. The asymptotically optimal solution has been found and involves the Kullback-Liebler information distance.

Currently, research is being carried out by Prof. J. Tsitsiklis and a graduate student, Mr. George Polychronopoulos, and involves the following two directions.

- (a) We have considered a class of symmetric detection problems in which given any hypothesis H_i , each sensor has probability ϵ of making an observation indicating that some other hypothesis H_j is true. A simple numerical procedure has been found which completely solves this problem. Furthermore, a closed form formula for the optimal decision rules has been found for the case where the "noise intensity" ϵ is very small.
- (b) In the context of the above symmetric problem we have posed problems of the following type: "Is it preferable to have N sensors each one transmitting D bits, or N/K sensors, each one transmitting KD bits? A complete solution has been found. The formulation represents a fundamental design problem in the design of distributed sensor systems.

We have also conducted research which addresses the issue of the validity of asymptotic considerations, when the number of agents N is moderate ($N=5$), with encouraging results.

The above results will be reported in the Masters thesis of Mr. Polychronopoulos (expected in the spring of 1988) and on a subsequent journal paper.

Documentation

- [1] J. N. Tsitsiklis, "On Threshold Rules in Decentralized Detection," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986; also LIDS-P-1570, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, June 1986.
- [2] J. N. Tsitsiklis, "Decentralized Detection by a Large Number of Sensors," LIDS-P-1662, April 1987; to appear in *Mathematics of Control, Signals and Systems*, 1988.

3.3 COMMUNICATION REQUIREMENTS OF DIVISIONALIZED ORGANIZATIONS

Background: In typical organizations, the overall performance cannot be evaluated simply in terms of the performance of each subdivision, as there may be nontrivial coupling effects between distinct subdivisions. These couplings have to be taken explicitly into account; one way of doing so is to assign to the decisionmaker associated with the operation of each division a cost function which reflects the coupling of his own division with the remaining divisions. Still, there is some freedom in such a procedure: For any two divisions A and B it may be the responsibility of either decisionmaker A or decisionmaker B to ensure that the interaction does not deteriorate the performance of the organization. Of course, the decisionmaker in charge of those interactions needs to be informed about the actions of the other decisionmaker. This leads to the following problem. Given a divisionalized organization and an associated organizational cost function, assign cost functions to each division of the organization so that the following two goals are met: a) the costs due to the interaction between different divisions are fully accounted for by the subcosts of each division; b) the communication interface requirements between different divisions are small.

In order to assess the communication requirements of a particular assignment of costs to divisions, we take the view that the decisionmakers may be modeled as boundedly rational

individuals, that their decisionmaking process consists of a sequence of adjustments of their decisions in a direction of decreasing costs, while exchanging their tentative decisions with other decisionmakers who have an interest in those decisions. We then require that there are enough communications so that this iterative process converges to an organizationally optimal set of decisions.

Problem Statement: Consider an organization with N divisions and an associated cost function $J(x_1, \dots, x_N)$, where x_i is the set of decisions taken at the i -th division. Alternatively, x_i may be viewed as the mode of operation of the i -th division. The objective is to have the organization operating at a set of decisions (x_1, \dots, x_N) which are globally optimal, in the sense that they minimize the organizational cost J . We associate with each division a decisionmaker DM_i , who is in charge of adjusting the decision variables x_i . We model the decisionmakers as "boundedly rational" individuals; mathematically, this is translated to the assumption that each decisionmaker will slowly and iteratively adjust his decisions in a direction which reduces the organizational costs. Furthermore, each decisionmaker does so based only on partial knowledge of the organizational cost, together with messages received from other decisionmakers.

Consider a partition $J(x_1, \dots, x_N) = \sum_{i=1}^N J^i(x_1, \dots, x_N)$ of the organizational cost. Each subcost J^i reflects the cost incurred to the i -th division and in principle should depend primarily on x_i and only on a few of the remaining x_j 's. We then postulate that the decisionmakers adjust their decisions by means of the following process (algorithm):

(a) DM_i keeps a vector x with his estimates of the current decision x_k of the other decisionmakers; also a vector λ with estimates of $\lambda^k_i = \partial J^k / \partial x_i$, for $k \neq i$. (Notice that this partial derivative may be interpreted as DM_i 's perception of how his decisions affect the costs incurred to the other divisions.

(b) Once in a while DM_i updates his decision using the rule $x_i := x_i + \gamma \sum_{k=1}^N \lambda^k_i$, (γ is a small

positive scalar) which is just the usual gradient algorithm.

- (c) Once in a while DM_i transmit his current decision to other decisionmakers.
- (d) Other decisionmakers reply to DM_i , by sending an updated value of the partial derivative $\partial J^k / \partial x_i$.

It is not hard to see that for the above procedure to work it is not necessary that all DM's communicate to each other. In particular, if the subcost J_i depends only on x_i , for i , there would be no need for any communication whatsoever. The required communications are in fact determined by the sparsity structure of the Hessian matrix of the subcost functions J_i . Recall now that all that is given is the original cost function J ; we therefore, have freedom in choosing the J_i 's and we should be able to do this in a way that introduces minimal communication requirements; that is, we want to minimize the number of pairs of decisionmakers who need to communicate to each other.

Progress to Date: A graduate student, C. Lee, supervised Prof. J. Tsitsiklis, undertook the task of formulating the problem of finding partitions that minimize the number of pairs of DM's who need to communicate to each other, as the topic of his SM research. It was realized that with a naive formulation the optimal allocation of responsibilities, imposing minimal communication requirements, corresponds to the centralization of authority. Thus, in order to obtain more realistic and meaningful problems we did incorporate a constraint requiring that no agent should be overloaded. A number of results have been obtained for a class of combinatorial problems, corresponding to the problem of optimal organizational design, under limited communications. In particular certain cases were solved; other cases have been successfully reformulated as linear network flow or assignment problems, for which efficient algorithms are known, and finally, some cases were shown to be intractable combinatorial problems (NP-complete).

This line of research is now essentially complete. Most results have been reported in the Masters thesis of Mr. C. Lee [1]. A journal paper will be prepared in the next few months covering both the philosophical and the technical aspects of this work.

Documentation:

- [1] C. Lee, "Task Allocation for Efficient Performance of a Decentralized Organization, LIDS-TH-1706, S.M. Thesis, Laboratory for Information and Decision Systems, MIT,

Cambridge, MA, September 1987.

3.4 COMMUNICATION COMPLEXITY IN DISTRIBUTED PROBLEM SOLVING

Background: The objective of this research effort is to quantify the minimal amount of information that has to be exchanged in an organization, subject to the requirement that a certain goal is accomplished, such as the minimization of an organizational cost function. The problem becomes interesting and relevant under the assumption that no member of the organization "knows" the entire function being minimized, but rather each agent has knowledge of only a piece of the cost function. A normative/prescriptive solution is sought.

Problem Formulation: Let f and g be convex function of n variables. Suppose that each one of two agents (or decisionmakers) knows the function f (respectively g), in the sense that he is able to compute instantly any quantities associated with this function. The two agents are to exchange a number of binary messages until they are able to determine a point x such that $f(x) + g(x)$ comes within ϵ of the minimum of $f+g$, where ϵ is some prespecified accuracy. The objective is to determine the minimum number of such messages that have to be exchanged, as a function of ϵ and to determine communication protocols which use no more messages than the minimum amount required.

Results: Several variations of this problem have been studied and solved by Professor J. Tsitsiklis and a graduate student Zhi-Quan Luo. Results have been reported in [1].

An interesting qualitative feature of the communication-optimal algorithms discovered thus far is the following: It is optimal to transmit aggregate information (the most significant bits of the gradient of the function optimized) in the beginning; then, as the optimum is approached more refined information should be transferred. This very intuitive result seems to correspond to realistic situations in human decisionmaking.

More recently, we have considered a new formulation in which the messages are real-valued, rather than discrete. A prototype problem is to assume that each one of two agents knows a $n \times n$ matrix A_i , $i = 1, 2$. The objective is to compute a particular entry of $(A_1 + A_2)^{-1}$. This problem arises, for example in distributed optimization of a cost function of the form $x'A_1x + x'A_2x + x'b$.

An obvious solution is for agent 1 to transmit all of the entries of A_1 to agent 2 who then performs the required computations. This scheme requires n^2 communications. We have succeeded in showing that there exists no method which will do with fewer than $O(n^2)$ communications. That is information must be centralized. On the technical side, we have restricted to communication protocols which are smooth rational functions of the original data A_1 , A_2 . (Otherwise n^2 numbers could have been coded in a single real number). The proof of our result uses novel techniques and makes use of certain results in algebraic geometry.

Documentation:

- [1] J. N. Tsitsiklis and Z.-Q. Luo, "Communication Complexity of Convex Optimization," LIDS-P-1617, Laboratory for Information and Decision Systems, MIT, October 1986; *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December, 1986; also an invited talk was given at the *2nd Symposium on Complexity of Approximately Solved Problems*, Columbia University, New York, April 1987; also, *Journal of Complexity*, 3, 1987, pp. 231-243.

3.5 DISTRIBUTED ORGANIZATION DESIGN

Background: The bounded rationality of human decisionmakers and the complexities of the tasks they must perform mandate the formation of organizations. Organizational architectures distribute the decisionmaking workload among the members: different architectures impose different individual loads and result in different organizational performance. Two measures of organizational performance are accuracy and timeliness. The first measure of performance addresses in part the quality of the organization's response. The second measure reflects the fact that in tactical decisionmaking *when* a response is generated is also significant: the ability of an organization to carry out tasks in a timely manner is a determinant factor of effectiveness.

The scope of work was divided into three tasks:

- (a) Evaluation of Alternative Organizational Architectures;
- (b) Asynchronous Protocols; and
- (c) Information Support Structures.

During this year, the research effort has been organized around three foci. In the first one, we

continue to work on the development of analytical and algorithmic tools for the analysis and design of organizations. In the second, we are integrating the results obtained thus far through the development of a workstation for the design and analysis of alternative organizational architectures. Finally, the experimental program, initiated last year with the objective of collecting data necessary to calibrate the models and evaluate different architectures for distributed decisionmaking, has been continuing and is expanding.

3.5.1 Design and Evaluation of Alternative Organizational Architectures.

In order to design an organization that meets some performance requirements, we need to be able to do the following:

- (a) Articulate the requirements in qualitative and quantitative terms;
- (b) Generate candidate architectures that meet some of the requirements;
- (c) Evaluate the candidate organizations with respect to the remaining requirements;
- (d) Modify the designs so as to improve the effectiveness of the organization;

The generalized Performance Workload locus has been used as the means for expressing both the requirements that the organization designer must meet and the performance characteristics of any specific design. Consider an organization with N decisionmakers. Then the Performance Workload space is an $N+2$ dimensional space in which two of the dimensions correspond to the measures of the organization's performance (say, accuracy and timeliness) and the remaining N dimensions correspond to the measure of the workload of each individual decisionmaker. Two loci can be defined. First, the Requirements locus is the set of points in this $N+2$ dimensional space that satisfy the performance and workload requirements associated with the task to be performed by the organization. The second, the System locus, is the set of points that are achievable by a particular design. The design problem can then be conceptualized as the reshaping and repositioning of the System locus in the Performance Workload space so that the requirements are met.

Three thesis projects were completed during this period. The individual problem statements and a description of the results follow:

Modeling and Evaluation of Variable Structure Organizations

Problem Statement: Develop a methodology for modeling and analyzing classes of variable-structure organizations, i.e., organizations where the interactions between decision makers can change. These organizations, named VDMO from now on, can be classified according to what factors trigger the change. Three types have been defined:

- Type 1 Variability: The VDMO adapts the structure of its interactions to the input it processes.
- Type 2 Variability: The VDMO adapts the structure of its interactions to changes in the environment in which it functions.
- Type 3 Variability: The VDMO adapts the structure of its interactions to changes in its own components. For instance, it can reconfigure itself to perform its task when its resource availability has changed.

In both Type 2 and 3 VDMOs, the issue of the detection by the organization that a change has occurred has not been addressed. These three types of variability can exist concurrently in a given organization; however, for their analysis and for the evaluation of their effects on system performance, they have been treated separately.

Progress to Date: This problem was addressed by Jean-Marc Monguillet under the supervision of Dr. A. H. Levis. The focus of the research effort has been the modeling and analysis of variable structure organizations using Predicate Transition Nets.

The System Effectiveness Analysis methodology has been extended to account for variable structure organizations. A Measure of Effectiveness has been proposed for each type of variable DMO. A mathematical formulation for the computation of that MOE has been established.

A modeling methodology has been described providing a representation of DMO's by functions. The main features of that methodology is the decoupling between the pattern of interactions and the identity of decisionmakers, who are modeled by tokens and treated like any other resources. The Predicate Transition Nets formalism has been adapted to allow such representation.

An example illustrating the overall procedure has been developed. It consists of three candidate designs for an air defense task. Each of these candidates is composed of three decisionmakers, namely one Headquarters and two Field Units. Two organizations have a fixed structure, and the

third one is type 1 variable; for some tasks, it adapts the pattern of interactions to a pattern comparable to that of the first fixed structure DMO. For some others, it takes the other pattern. The results of the comparison of these designs are that a particular organizational design cannot be selected in general on the basis of its performance characteristics alone, as presented in the form of a system locus. The Effectiveness of each candidate has to be evaluated quantitatively for each set of mission requirements; then zones can be defined in the requirements space which characterize for each organization the ranges of mission requirements for which it is the most effective. In that particular case, the set of mission requirements for which the variable structure organization has the highest Effectiveness can be computed. It has been shown clearly that a variable structure organization was preferable to the fixed structure ones when the requirements were such that one fixed design was not timely enough, whereas the other was not accurate enough. Type 1 variability provided a compromise between extreme performance of organizations with fixed structure.

Documentation: The thesis of J.-M. Monguillet has been issued as a LIDS report.

- [1] Jean-Marc Monguillet, "Modeling and Evaluation of Variable structure Organizations," S.M. Thesis, Report LIDS-TH-1730, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, December 1987.

Design of Organizations

Objective: Given a feasible organizational architecture, develop a methodology for (a) identifying the functions that must be performed by the organization in order that the task be accomplished, (b) selecting the resources (human, hardware, software) that are required to implement these functions, and (c) integrating these resources - through interactions - so that the system operates effectively.

Progress to Date: This research problem has been investigated by Stamos K. Andreadakis under the supervision of Dr. A. H. Levis. A doctoral thesis has been defended successfully in December and the dissertation is in the final stages of preparation. The results of this task will be reported in the next progress report.

Documentation:

- [1] A. H. Levis and S.K. Andreadakis, "Computer-Aided Analysis of Organizations," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.
- [2] S.K. Andreadakis and A. H. Levis, "Accuracy and Timeliness in Decision-Making Organizations," *Proc. 10th IFAC World Congress*, July 27-31, 1987, Munich, FRG and *Proc. 9th MIT/ONR Workshop on C³ Systems*, LIDS-R-1624, MIT Cambridge, MA, December 1986.
- [3] S. K. Andreadakis and A. H. Levis, "Design Methodology for Decision-Making Organizations," *Proc. of 1987 Symposium on C² Research*, National Defense University, Washington DC, June 1987.

Performance Evaluation of Organizations with Decision Aids

Problem Statement: Analyze and evaluate the impact of decision aids, i.e., preprocessors and decision support systems, on the effectiveness of decisionmaking and information processing organizations. In particular, investigate the concept of coordination of decisionmakers assisted by those decision aids.

Progress to Date: A Master's Thesis has been completed by Jean-Louis Grevet under the supervision of Dr. A.H. Levis. From a conceptual standpoint, the idea of coordination in decision-making organizations embodies three classes of issues:

- the extent to which the decisionmakers constitute a team.
- the synchronization of the decisionmakers' activities .
- the consistency of the information processed by the different members of the organization.

The latter class of issues is primarily related to the fact that decisionmakers do not necessarily process data that are consistent because they have different geographical or temporal origins : For instance, two different decisionmakers can process data originating from different sensors or different databases as well as data originating from a common database but accessed at different instants.

The work focused on these three of issues:

(a) The concept of team of decision-makers has been clarified. A team of decisionmakers is defined as being an organization in which the members :

- have a common goal
- have the same interests and same perception of the environment
- have activities which must be coordinated so that they achieve a higher performance.

Thus, for a task X with probability distribution $p(X)$ and a cost function $c(X)$ for the organization, one condition for the organization to be a team is that its members have the same perception of the task, $p_T(X)$, i.e. the same beliefs about the task, and assign the same cost $c_T(X)$ to each input, i.e., have the same interests as far as the task is concerned.

The team will account perfectly for the organizational objectives when:

$$p_T(X) = p(X) \quad \text{and} \quad c_T(X) = c(X)$$

(b) The issue of synchronization is related to the interactions between the decisionmakers that take place during the decisionmaking process. It is thus a dynamic characteristic of the organization. When a decisionmaker DM_i processes some information, the total processing time of this input for DM_i consists of two distinct parts:

- the time T_i during which the decisionmaker actually processes the information
- the time T_p spent by the information in the memory of the decisionmaker without being processed.

The time T_p is the result of two factors:

- information can remain in the memory of the decisionmaker until he decides to process it using the relevant algorithm. In this case, the decisionmaker processes several pieces of information at the same time. Since a particular algorithm cannot process two inputs at the same time, some inputs will have to remain unprocessed in memory waiting until the relevant algorithm is free.
- Information can also stay in memory because the decisionmaker waits to receive a

necessary piece of information from another decisionmaker or a decision support system.

An organization is perfectly synchronized when, for the whole decisionmaking process, the decisionmakers do not have to wait for information that they need in order to process the information that is in memory. The synchronization degrades when the processing of some inputs leads decisionmakers to wait for these data.

Synchronization is an important concept because the processing of information introduces three kinds of biases:

- biases due to the uncertainty embodied in the information processe;
- biases due to the models used; and
- biases due to the value of information when the decisionmaker actually processes it.

If an item of information remains in memory for along time, the decisionmaker might well attach less value to it when he actually processes it. This could lead to a degradation of the effectiveness of the organization.

- (b) The consistency of information refers to whether or not different items of data can be fused together without contraction. It is mission dependent. Data can be inconsistent if they have different geographical or temporal origins: For instance, two different decisionmakers can process data originating from different sensors or different databases as well as data originating from a common database but accessed at different times.

The modeling of decisionmaking processes that require coordination has been completed using the Predicate Transition Nets formalism. The tokens, which are the symbolic information carrier, are identified by three attributes:

- the time of entry in the net, T_n .
- the time of entry in a specific place, T_d .
- the class c assigned to information items by the previous processing stage.

The rule of enablement of transitions is that the tokens in the input places must have the same

attribute T_n . It means that, when decisionmakers interact, they must refer to the same input.

Two measures that can be used for evaluating the coordination in decisionmaking organizations have been defined:

- the degree of information consistency, D .
- the measure of synchronization, S_T .

A simulation program of Predicate Transition Nets have been developed using the Design Open Architecture Development Systems of Meta Software Corp., has been developed. It can be used to get insight in the dynamics of decisionmaking process.

The impact of decision-aids on the coordination of decisionmaking organizations has been assessed using the modeling and evaluation tools described above. A model of a decisionmaker assisted by a decision support system has been proposed. It accounts for the fact that most real systems contain both elements of centralization and decentralization, i.e., the users can share certain resources - centralized databases or mainframes - and access individually other facilities such as intelligent terminals. This modifies the strategy of each decisionmaker who must integrate in his choices the possibility of requesting information from the DSS. Thus, each decisionmaker has three alternatives vis-a-vis the DSS:

- he can ignore it and process the information by himself.
- he can query it and rely totally on the response.
- he can query it and compare the response to his own perception of the issue.

The evaluation of these choices has been carried out on an example, a two-person hierarchical organization.

It has been found that decision-aids can modify the coordination of decisionmaking process by:

- modifying a priority order with which different organization members process the inputs.
- increasing the number of information flow paths with different processing times.

Documentation:

- [1] J. L. Grevet, "Decision Aiding and Coordination Decision-Making Organizations," MS Thesis, LIDS-TH-1737, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, January 1988.
- [2] J. L. Grevet and A. H. Levis, "Coordination in Decision-Making Organizations," LIDS-P-1738, also to appear in *Proc. 9th European Workshop on Applications and Theory of Petri Nets*, Venice, Italy, June 22-24, 1988.
- [3] J. L. Grevet, L. Jandura, J. Brode, and A. H. Levis, "Execution Strategies for Petri Net Simulation," to appear in *Proc. 12th IMACS World Congress on Scientific Computation*, Paris, France, July 18-22, 1988.

MEETINGS**October 1, 1987: Newport, R.I.**

Professors Athans and Tsitsiklis, Dr. Levis and research assistants, S. K. Andreadakis, J. Azzola, V. Jin, J. Kyrtzoglou and J. Papastavrou attended the Annual Review of the DTDM program organized by the Office of Naval Research.

October 29, 1987: Kansas City, KS

Dr. Levis attended the 5th Annual Workshop on Command and Control Decision Aiding where he presented one paper and participated in a panel discussion.

November 15, 1987: Oxnard, CA

Dr. Levis attended the meeting of the DOD Technical Groups on Human Factors Engineering and Man-Machine Modeling and delivered an invited paper.

November 25, 1987: San Diego, CA (NOSC)

Dr. Levis attended the Annual Review meeting of the Technical Panel on C³ of the Joint Directors of Laboratories and presented a review of research on C³ at MIT/LIDS. He also discussed MIT

work and its relevance for ASW problems with the ASW group at NOSC.

December 8, 1987: Los Angeles, CA

Dr. Levis attended the 26th IEEE Conference on Decision and Control, and participated in a panel on SDI/Battle Management.

5. RESEARCH PERSONNEL

Prof. Michael Athans,	Co-principal investigator
Dr. Alexander H. Levis,	Co-principal investigator
Prof. John Tsitsiklis	
Mr. Stamatios Andreadakis	graduate research assistant (Ph.D)
Ms. Victoria Jin	graduate research assistant (Ph.D)
Mr. Jason Papastavrou	graduate research assistant (Ph.D)
Mr. Jean-Louis Grevet	graduate research assistant (M.S.)
Mr. Jean-Marc Monguillet	graduate research assistant (M.S.)
Ms. Anne-Claire Louvet	graduate research assistant (M.S.)
Mr. Javid Pothiwala	graduate research assistant (M.S.)

6. DOCUMENTATION

6.1 Theses

- [1] J. N. Tsitsiklis, "Problems in Decentralized Decisionmaking and Computation," Ph.D. Thesis, Dept. of EECS also, Report LIDS-TH-1424, Laboratory for Information and Decision Systems, MIT, December 1984.
- [2] G. J. Bejjani, "Information Storage and Access in Decisionmaking Organizations," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1434, Laboratory for Information and Decision Systems, MIT, January 1985.
- [3] Y. V.-Y. Jin, "Delays for Distributed Decisionmaking Organizations," MS Thesis,

Mechanical Engineering, also, Report LIDS-TH-1459, Laboratory for Information and Decision Systems, MIT, May 1985.

- [4] K. L. Boettcher, "A Methodology for the Analysis and Design of Human Information Processing Organizations," Ph.D., Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH-1501, Laboratory for Information and Decision Systems, MIT, September 1985.
- [5] R. P. Wiley, "Performance Analysis of Stochastic Timed Petri Nets," Ph.D. Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH-1525, Laboratory for Information and Decision Systems, MIT, January 1986.
- [6] J. D. Papastavrou, "Distributed Detection with Selective Communications," MS Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH-1563, Laboratory for Information and Decision Systems, MIT, May 1986.
- [7] C.M. Bohner, "Computer Graphics for Systems Effectiveness Analysis," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1573, Laboratory for Information and Decision Systems, MIT, June 1986.
- [8] H. P. Hillion, "Performance Evaluation of Decisionmaking Organizations Using Timed Petri Nets," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1590, MIT, August 1986.
- [9] P. A. Remy, "On the Generation of Organizational Architectures Using Petri Nets," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1630, Laboratory for Information and Decision Systems, MIT, December 1986.
- [10] C. Lee, "Task Allocation for Efficient Performance of a Decentralized Organization," MS, Thesis, Dept of EECS, also, Report LIDS-TH-1706, Laboratory for Information and Decision Systems, MIT, September 1987.
- [11] J. M. Monguillet, "Modeling and Evaluation of Variable Structure Organizations," MS Thesis, LIDS-TH-1730, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, December 1988.
- [12] J. L. Grevet, "Decision Aiding and Coordination Decision-Making Organizations," MS Thesis, LIDS-TH-1737, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, January 1988.

6.2 Technical Papers

- [1] R. P. Wiley and R. R. Tenney, "Performance Evaluation of Stochastic Timed Decision-Free Petri Nets," *Proc. 24th IEEE Conference on Decision and Control*, December 11-13, 1985, Ft. Lauderdale, FL, pp. 558-563.
- [2] G. J. Bejjani and A. H. Levis, "Information Storage and Access in Decisionmaking Organizations," LIDS-P-1466, Laboratory for Information and Decision Systems, MIT, May 1986.
- [3] J. N. Tsitsiklis and M. Athans, "On the Complexity of Decentralized Decisionmaking

- and Detection Problems, *IEEE Trans. on Automatic Control*, Vol. AC-30, No. 5, May 1985.
- [4] V. Y.-Y. Jin and A. H. Levis, "Computation of Delays in Acyclical Distributed Decisionmaking Organizations," LIDS-P-1488, Laboratory for Information and Decision Systems, MIT, August 1985.
 - [5] K. L. Boettcher, and R. R. Tenney, "A Case Study in Human Team Decisionmaking," *Proc. Systems, Man, and Cybernetics*, Tucson, AZ, November 1985.
 - [6] K. L. Boettcher and R. R. Tenney, "On the Analysis and Design of Human Information Processing Organizations," LIDS-P-1503, *Proc. 8th MIT/ONR Workshop on C³ Systems*, December 1985.
 - [7] K. L. Boettcher and R. R. Tenney, "Distributed Decisionmaking with Constrained Decisionmakers - A Case Study," *Proc. 8th MIT/ONR Workshop on C³ Systems*, December 1985, also in *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-16, No. 6, November/December 1986.
 - [8] R. P. Wiley and R. R. Tenney, "Calculating Timed-Related Performance Measures of a Distributed Tactical Decisionmaking Organization Using Stochastic Timed Petri Nets," *Proc. 8th MIT/ONR Workshop on C³ Systems*, December 1985.
 - [9] V. Y.-Y. Jin, A. H. Levis, and P. A. Remy, "Delays in Acyclical Distributed Decisionmaking Organizations," LIDS-P-1528, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, also in *Proc. 4th IFAC/IFORS Symposium on Large Scale Systems: Theory and Applications*, Zurich, Switzerland, August 1986.
 - [10] J. Papastravrou and M. Athans, "A Distributed Hypotheses-Testing Team Decision Problem with Communications Cost," LIDS-P-1538, Laboratory for Information and Decision Systems, MIT, February 1986, also in *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.
 - [11] J. Tsitsiklis, "On Optimal Thresholds in Decentralized Detection," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.
 - [12] A. H. Levis and S. K. Andreadakis, "Computer-Aided Design of Organizations," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.
 - [13] P. Remy, A. H. Levis, V. Jin, "On the Design of Distributed Organizational Structures," LIDS-P-1581, Laboratory for Information and Decision Systems, MIT, July 1986, *Proc. 10th IFAC World Congress*, Munich, FRG, July 1987.
 - [14] M. Athans, "Command-and-Control (C²) Theory: A Challenge to Control Science," *IEEE Trans. on Automatic Control*, Vol. AC-32, No. 4, April 1987, pp. 286-293.
 - [15] P.A. Remy and A. H. Levis, "On the Generation of Organizational Architectures Using Petri Nets," *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987.
 - [16] H.P. Hillion and A.H. Levis, "Timed Event-Graph and Performance Evaluation of

Systems, *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987.

- [17] J. N. Tsitsiklis, D.P. Bertsekas, and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Trans. on Automatic Control*, Vol. AC-31, No. 9, September 1986.
- [18] S. K. Andreadakis and A. H. Levis, "Design Methodology for Decision-Making Organizations," *Proc. 1987 C³ Symposium*, National Defense University, Fort McNair, Washington DC, June 1987.
- [19] A. C. Louvet, J. T. Casey, A. H. Levis, "Experimental Investigation of the Bounded Rationality Constraint," *Proc. 1987 C³ Symposium*, National Defense University, Fort McNair, Washington DC, June 1987.
- [20] H. P. Hillion and A. H. Levis, "Performance Evaluation of Decisionmaking Organizations," *Proc. 1987 C³ Symposium*, National Defense University, Fort McNair, Washington DC, June 1987.
- [21] S. T. Weingaertner and A. H. Levis, "Evaluation of Decision Aiding in Submarine Emergency Decisionmaking," *Proc. 1987 C³ Symposium*, National Defense University, Fort McNair, Washington DC, June 1987.
- [22] P. J. F. Martin and A. H. Levis, "Measures of Effectiveness and C³ Testbed Experiments," *Proc. 1987 C³ Symposium*, National Defense University, Fort McNair, Washington DC, June 1987.
- [23] P. A. Remy and A. H. Levis, "On the Generation of Organizational Architectures Using Petri Nets," *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987; to appear in *Advances in Petri Nets*, 1988, Springer-Verlag, Berlin.
- [24] H. P. Hillion and A. H. Levis, "Timed Event-Graph and Performance Evaluation of Systems," *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987.
- [25] J. Papastavrou and M. Athans, "A Distributed Hypothesis-Testing Team Decision Problem with Communications Cost," in *S.Tzafestas et al. (eds.), System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, Volume 1, pp. 99-130, 1987, D. Reidel Publishing Company.
- [26] J. N. Tsitsiklis and Z.-Q. Lee, "Communication Complexity of Convex Optimization," *Journal of Complexity*, 3, pp. 231-243, 1987.
- [27] P.A. Remy, A. H. Levis, and V. Y.-Y. Jin, "On the Design of Distributed Organizational Structures," *Automatica*, January 1988.
- [28] P. Remy and A. H. Levis, "On the Generation of Organizational Architectures Using Petri Nets," to appear in *Proc. 12th IMACS World Congress on Scientific Computation*, Paris, France, July 1988.
- [29] J. L. Grevet and A. H. Levis, "Coordination in Decision-Making Organizations,"

LIDS-P-1738, also to appear in *Proc. 9th European Workshop on Applications and Theory of Petri Nets*, Venice, Italy, June, 1988.

- [30] J. L. Grevet, L. Jandura, J. Brode, and A. H. Levis, "Execution Strategies for Petri Net Simulation," to appear in *Proc. 12th IMACS World Congress on Scientific Computation*, Paris, France, July, 1988.

Communication Complexity of Convex Optimization*

JOHN N. TSITSIKLIS AND ZHI-QUAN LUO

*Laboratory for Information and Decision Systems and the Operations Research Center,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

We consider a situation where each of two processors has access to a different convex function f_i , $i = 1, 2$, defined on a common bounded domain. The processors are to exchange a number of binary messages, according to some protocol, until they find a point in the domain at which $f_1 + f_2$ is minimized, within some ϵ -specified accuracy. Our objective is to determine protocols under which the number of exchanged messages is minimized. © 1987 by Academic Press, Inc.

1. INTRODUCTION

Let $\tilde{\mathcal{F}}$ be a set of convex functions defined on the n -dimensional bounded domain $[0, 1]^n$. (Typically, $\tilde{\mathcal{F}}$ will be defined by imposing certain smoothness conditions on its elements.) Given any $\epsilon > 0$, and $f \in \tilde{\mathcal{F}}$, let $R(f; \epsilon)$ be the set of all $x \in [0, 1]^n$ such that $f(x) \leq f(y) + \epsilon$, $\forall y \in [0, 1]^n$.

Let there be two processors, denoted by P_1 and P_2 . Each processor is given a function $f_i \in \tilde{\mathcal{F}}$. Then they start exchanging binary messages, according to some protocol π , until processor P_1 determines an element of $R(f_1 + f_2; \epsilon)$. Let $C(f_1, f_2; \epsilon, \pi)$ be the total number of messages that are exchanged; this is a function of the particular protocol being employed and we are looking for an optimal one. More precisely, let

$$C(\tilde{\mathcal{F}}; \epsilon, \pi) = \sup_{f_1, f_2 \in \tilde{\mathcal{F}}} C(f_1, f_2; \epsilon, \pi) \quad (1.1)$$

* Research supported by the Army Research Office under Contract DAAAG-29-84-K-0005 and by the Office of Naval Research under Contract N00014-84-K-0519 (NR 649-003). A preliminary version of this paper was presented at the 25th IEEE Conference on Decision and Control, Athens, Greece, Dec. 1986.

be the communication requirement (in the worst case) of the particular protocol and let

$$C(\bar{\mathcal{F}}; \epsilon) = \inf_{\pi \in \Pi(\epsilon)} C(\bar{\mathcal{F}}; \epsilon, \pi) \quad (1.2)$$

be the communication requirement under an optimal protocol, where $\Pi(\epsilon)$ is the class of all protocols which work properly, for a particular choice of ϵ . The quantity $C(\bar{\mathcal{F}}; \epsilon)$ may be called the ϵ -communication complexity of the above-defined problem of distributed, approximate, convex optimization.

For the above definition to be precise, we need to be specific regarding the notion of a protocol; that is, we have to specify the set $\Pi(\epsilon)$ of admissible protocols and this is what we do next. A protocol π consists of

- (a) A termination time T ;
- (b) A collection of functions $M_i : \bar{\mathcal{F}} \times \{0, 1\}^i \rightarrow \{0, 1\}$, $i = 1, 2, \dots, T-1$;
- (c) A final function $Q : \bar{\mathcal{F}} \times \{0, 1\}^T \rightarrow [0, 1]^n$.

A protocol corresponds to the following sequence of events. Each processor P_i receives its "input" f_i and then, at each time t , transmits to the other processor P_j a binary message $m_i(t)$ determined by

$$m_i(t) = M_i(f_i, m_i(0), \dots, m_i(t-1)).$$

Thus the message transmitted by a processor depends only on the function f_i known by it, together with all messages it has received in the past. At time T the exchange of messages ceases and processor P_1 picks a point in $[0, 1]^n$ according to

$$x = Q(f_1, m_2(0), \dots, m_2(T-1)). \quad (1.3)$$

The number $C(f_1, f_2; \epsilon, \pi)$ of messages transmitted under this protocol is simply $2T$. We define $\Pi(\epsilon)$ as the set of all protocols with the property that the point x generated by (1.3) belongs to $I(f_1 + f_2; \epsilon)$, for every $f_1, f_2 \in \bar{\mathcal{F}}$.

A couple of remarks on our definition of protocols are in order.

(i) We have constrained each processor to transmit exactly one binary message at each stage. This may be wasteful if, for example, a better protocol may be found in which P_1 first sends many messages and then P_2 transmit its own messages. Nevertheless, the waste that results can be at most a factor of two. Since, in this paper, we study only orders of magnitude, this issue is unimportant.

(ii) We have assumed that the termination time T is the same for all f_1 .

$f_i \in \mathcal{F}$, even though for certain "easy" functions the desired result may have been obtained earlier. Again, this is of no concern because we are interested in a worst case analysis.

Related Research

The study of communication complexity was initiated by Abelson (1980) and Yao (1979). Abelson deals with problems of continuous variables, in which an exact result is sought, and allows the messages to be real-valued, subject to a constraint that they are smooth functions of the input. This is a different type of problem from ours, because we are interested in an approximate result and we are assuming binary messages.

Yao (1979) deals with combinatorial problems, in which messages are binary and an exact result is obtained after finitely many stages. This reference has been followed by a substantial amount of research which developed the theory further and also evaluated the communication complexity of selected combinatorial problems (Papadimitriou and Sipser, 1982; Papadimitriou and Tsitsiklis, 1982; Aho *et al.*, 1983; Pang and El Gamal, 1986; Mehlhorn and Schmidt, 1982; Ullman, 1984). The main application of this research has been in VLSI, where communication complexity constrains the amount of information that has to flow from one side of a chip to the other; this in turn determines certain trade-offs on the achievable performance of special-purpose VLSI chips for computing certain functions (Ullman, 1984).

Finally, communication complexity has been also studied for models of asynchronous distributed computation, in which messages may reach their destination after an arbitrary delay (Awerbuch and Gallager, 1985).

The communication complexity of the approximate solution of problems of continuous variables has not been studied before, to the best of our knowledge. However, there exists a large amount of theory on the information requirements for solving (approximately) certain problems such as nonlinear optimization, and numerical integration of differential equations (Nemirovsky and Yudin, 1983; Traub and Woźniakowski, 1980) ("information based complexity"). Here one raises questions such as, How many gradient evaluations are required for an algorithm to find a point which minimizes a convex function within some prespecified accuracy ϵ ? We can see that, in this type of research, information flows one way—from a "memory unit" (which knows the function being minimized) to the processor—and this is what makes it different from ours.

Outline

In Section II we establish straightforward lower bounds such as $C(\mathcal{F}; \epsilon) \geq O(n \log(1/\epsilon))$. In Section III we show that the naive distributed version of ellipsoid-type algorithms leads to protocols with $O(n^2 \log(1/\epsilon)(\log n +$

$\log(1/\epsilon)$ communication requirements and we show that this upper bound cannot be improved substantially within a restricted class of protocols. In Sections IV and V we partially close the gap between the above-mentioned upper and lower bounds by presenting protocols with $O(\log(1/\epsilon))$ communication requirements for the case $n = 1$ (Section IV) and with $O(n \log n - \log(1/\epsilon))$ communication requirements for the case of general n (Section V), under certain regularity assumptions on the elements of \mathcal{F} . In Section VI, we provide some discussion of possible extensions and questions which remain open.

II. LOWER BOUNDS ON $C(\mathcal{F}; \epsilon)$

Before we prove any lower bounds we start with a fairly trivial lemma whose proof is omitted.

LEMMA 2.1. *If $\mathcal{F} \subset \mathcal{G}$ then $C(\mathcal{F}; \epsilon) \leq C(\mathcal{G}; \epsilon)$.*

Let \mathcal{F}_Q be the set of quadratic functions of the form $f(x) = \|x - x^*\|^2$, with $x^* \in [0, 1]^n$ and where $\|\cdot\|$ is the Euclidean norm. Also, let \mathcal{F}_{∞} be the set of functions of the form $f(x) = \{\max_{i=1, \dots, n} |x - x_i^*|\}$, where $|x_i^*| \leq 1, \forall i$.

PROPOSITION 2.2. (i) $C(\mathcal{F}_Q; \epsilon) \geq O(n(\log n - \log(1/\epsilon)))$;

(ii) $C(\mathcal{F}_{\infty}; \epsilon) \geq O(n \log(1/\epsilon))$.

Proof. (i) Consider a protocol $\pi \in \Pi(\epsilon)$ with termination time T and let us study its operation for the special case where $f_1 = 0$. Let S be the range of the function Q corresponding to that protocol (see Eq. (1.3)), when $f_1 = 0$. Given that the minimum of f_2 may be anywhere in $[0, 1]^n$, S must contain points which come within $\epsilon^{1/2}$, in Euclidean distance, from every point in $[0, 1]^n$. Now, one needs at least $(An/\epsilon^{1/2})^{Bn}$ Euclidean balls of radius $\epsilon^{1/2}$ to cover $[0, 1]^n$, where A and B are absolute constants. (This follows by simply taking into account the volume of a ball in n -dimensional space.) Therefore, the cardinality of S is at least $(An/\epsilon^{1/2})^{Bn}$. Given that the cardinality of the range of a function is no larger than the cardinality of its domain, it follows that the cardinality of S is no larger than 2^T . Therefore, $T \geq O(n(\log n - \log(1/\epsilon)))$, which proves the first part.

(ii) The proof is almost identical to that of part (i) and is therefore omitted. The only difference is that now $[0, 1]^n$ is covered by balls in the supremum norm and $O((1/\epsilon)^n)$ such balls are necessary and sufficient. ■

In the proof of Proposition 2.2 we made use of the assumption that the final result is always obtained by processor P_1 . Nevertheless, at the cost of minor complications of the proof, the same lower bound may be obtained even if we enlarge the class of allowed protocols so that the processor who computes the final result is not prespecified.

Let $\tilde{\mathcal{F}}_{SC, n, L}$ ("strongly convex functions") be the set of all continuously differentiable convex functions f with the properties

$$L\|x - y\|^2 \leq (f'(x) - f'(y))(x - y) \leq ML\|x - y\|^2, \quad (2.1)$$

$$f'(x) \leq MLn^{1/2}, \quad \forall x \in [0, 1]^n. \quad (2.2)$$

(Note that (2.1) implies that $M \geq 1$.) Also, let $\tilde{\mathcal{F}}_L$ be the set of convex functions which are bounded by $\frac{1}{4}$ and satisfy

$$f(x) - f(y) \leq \frac{1}{4} \max\|x - y\|, \quad \forall x, y.$$

PROPOSITION 2.3. (i) $C(\tilde{\mathcal{F}}_{SC, n, L}; \epsilon) \geq O(n \log n + \log(1/\epsilon))$.

(ii) $C(\tilde{\mathcal{F}}_L; \epsilon) \geq O(n \log(1/\epsilon))$.

Proof. Part (ii) follows from Proposition 2.2 and Lemma 2.1, because $\tilde{\mathcal{F}}_L \subseteq \tilde{\mathcal{F}}_1$. For part (i), we note that $\tilde{\mathcal{F}}_L \subseteq \tilde{\mathcal{F}}_{SC, n, 2}$ and Lemma 2.1 proves the result for $\tilde{\mathcal{F}}_{SC, n, 2}$. The result for general L follows because any $f \in \tilde{\mathcal{F}}_{SC, n, L}$ can be scaled so that it belongs to $\tilde{\mathcal{F}}_{SC, n, 2}$. ■

III. NAIVE UPPER BOUNDS

We consider here a straightforward distributed version of the method of the centers of gravity (MCG), which has been shown by Nemirovsky and Yudin (1983) to be an optimal algorithm in the single-processor case, for functions in $\tilde{\mathcal{F}}_L$, in the sense that it requires a minimal number of gradient evaluations. This method may be viewed as a generalization of the well-known ellipsoid algorithm for linear programming (Papadimitriou and Steiglitz, 1982). We start by describing the uniprocessor version of this method and then analyze the communication requirements of a distributed implementation.

The MCG Algorithm (Nemirovsky and Yudin, 1983, p. 62)

Let $f \in \tilde{\mathcal{F}}_L$ be a convex function to be minimized with accuracy ϵ . Let $G_0 = [0, 1]^n$ and let x_0 be its center of gravity. At the beginning of the k th stage of the computation, we assume that we are given a convex set $G_{k-1} \subset [0, 1]^n$ and its center of gravity x_k . Let z_k be a scalar and let y_k be a vector in R^n with the following properties:

$$(i) \quad z_k + \langle y_k, x - x_k \rangle \leq f(x), \quad \forall x \in [0, 1]^n;$$

$$(ii) \quad z_k \geq f(x_k) - (\epsilon/2).$$

(Note that if the term $\epsilon/2$ were absent in condition (ii), we would have $z_k =$

$f(x_k)$ and $y_k = f'(x_k)$, if x_k is an interior point. The presence of the $\varepsilon/2$ term implies that these relations need to hold only approximately.)

Let $a_k = \min_{x \in G_k} \{z\}$ and let $G_k = \{x \in G_{k-1} : (y_k, x - x_k) + z_k \leq a_k\}$. The algorithm terminates when the Lebesgue volume of G_k becomes smaller than $(\varepsilon/2)^n$ and returns a point x_k associated with the smallest value of z encountered so far.

The following facts are quoted from Nemirovsky and Yudin (1983).

- (a) The volume of G_k is no larger than a^k , where a is an absolute constant, smaller than one and independent of the dimension n . Thus a total of $n \log(2/\varepsilon) \log(1/a) = O(n \log(1/\varepsilon))$ stages are sufficient.
- (b) The result x of the algorithm satisfies $f(x) \leq \inf_{x \in \mathcal{F}_T} f(x) + V(x)$, where $V(x) = \sup_{f \in \mathcal{F}_T} f(x) - \inf_{f \in \mathcal{F}_T} f(x)$.

Note that $V(f) \leq 1$, for $f = f_1 + f_2$, $f_1, f_2 \in \mathcal{F}_T$ so that the algorithm indeed produces a result belonging to $I(f; \varepsilon)$.

We now consider a distributed implementation of this algorithm. The distributed protocol will consist of stages corresponding to the stages of the MCG algorithm. At the beginning of the k th stage, both processors know the current convex set G_{k-1} and are therefore able to compute its center of gravity x_k . Processor P evaluates $f(x_k)$ and transmits the binary representation of a message $b(i, k)$ satisfying $b(i, k) \in [f(x_k) - (\varepsilon/4), f(x_k) + (\varepsilon/8)]$. Clearly, $b(i, k)$ may be chosen so that its binary representation has at most $O(\log(1/\varepsilon))$ bits. Also, each processor evaluates the gradient $g_{i,k}$ of its function f_i at x_k (with components $g_{i,k,j}$, $j = 1, \dots, n$) and transmits the binary representation of messages $c(i, k, j)$ satisfying $|g_{i,k,j} - c(i, k, j)| \leq \varepsilon/(16n)$. Clearly the $c(i, k, j)$'s may be chosen so that they can be all transmitted using $O(n \log(n/\varepsilon)) = O(n \log n + n \log(1/\varepsilon))$ bits.

Next, each processor lets $z_i = b(1, k) - b(2, k)$ and lets y_k be the vector with components $c(1, k, j) - c(2, k, j)$. It then follows by some simple algebra that z_k and y_k satisfy the specifications of the MCG algorithm. Finally, each processor determines G_k and its center of gravity x_{k+1} , and the algorithm proceeds to its next stage.

We now combine our estimates of the number of stages of the MCG algorithm and of the communication requirements per stage to conclude the following.

PROPOSITION 3.1. $C(\mathcal{F}_T; \varepsilon) \leq O(n^2 \log(1/\varepsilon) \log n + \log(1/\varepsilon))$. In particular, the above-described distributed version of the MCG algorithm stays within this bound.

The upper bound of Proposition 3.1 is quite far from the lower bound of Proposition 2.2. We show next that within a certain class of protocols this upper bound cannot be substantially improved.

We consider protocols which consist of stages. At the k th stage there is a current point $x_k \in [0, 1]^n$ known by both processors. Then, the processors transmit to each other approximate values of f and of the gradient of

f_i , all evaluated at x_k . Using the values of these messages, together with any past common information, they determine the next point x_{k+1} , according to some commonly known rule, and so on. We place one additional restriction: when a processor transmits an approximate value of $f_i(x_k)$ it does so by transmitting a sequence of bits of the binary representation of $f_i(x_k)$ starting from the most significant one and continuing with consecutive less significant bits. (So, for example, a processor is not allowed to transmit the first and the third most significant bits of $f_i(x_k)$, without transmitting the second most significant bit.) The same assumption is made concerning the components of the gradient of f_i . Finally, we require that the same number of bits of $f_i(x_k)$ and of each component of the gradient of f_i get transmitted.

The above restrictions turn out to be quite severe.

PROPOSITION 3.2. *There exists a constant A such that for any protocol $\pi \in \Pi(\epsilon)$ satisfying the above restrictions, there exist $f_1, f_2 \in \mathcal{F}_1$ such that $C(f_1, f_2; \epsilon, \pi) \geq An^2 \log^2(1/\epsilon)$. This is true, even if we restrict f_1 to be equal to the identically zero function.*

Proof. Using an argument similar to Lemma 2.1, it is sufficient to prove the result under the restriction that $f_1 = 0$ and under the restriction that f_2 be differentiable and bounded, together with every component of its derivative, by $\epsilon^{1/2}$. Using the results of Nemirovsky and Yudin (1983), for processor P_1 to determine a point which is optimal within ϵ , it must acquire nontrivial information on the values and the derivatives of f_2 for at least $An \log(1/\epsilon^{1/2})$ different points. Note that the $O(\log(1/\epsilon^{1/2}))$ most significant bits of f_2 and each component of its derivative, evaluated at any point, are always zero. Thus, for processor P_1 to obtain nontrivial information at a certain point at least $O(n \log(1/\epsilon^{1/2}))$ bits have to be transmitted. This leads to a total communication requirement of $O(n^2 \log^2(1/\epsilon^{1/2})) = O(n^2 \log^2(1/\epsilon))$ bits, which proves the result. ■

If we relax the requirement that the same number of bits be transmitted for each component of the gradient, at each stage, then the same proof yields the lower bound $C(f_1, f_2; \epsilon, \pi) \geq An \log^2(1/\epsilon)$.

IV. AN OPTIMAL ALGORITHM FOR THE ONE-DIMENSIONAL CASE

We prove here a result which closes the gap between upper and lower bounds for the one-dimensional case. The proof consists of the construction of an optimal protocol. We only present the protocol under the assumption that each f_i is differentiable. The argument is the same in the nondifferentiable case, except that each f'_i is to be interpreted as a subgradient.

PROPOSITION 4.1. *If $n = 1$ then $C(\mathcal{F}_1; \epsilon) \leq O(\log(1/\epsilon))$.*

Proof. The protocol consists of consecutive stages. At the beginning of the k th stage, both processors have knowledge of four numbers, a_k , b_k , c_k , and d_k , with the following properties:

- (i) The interval $[a_k, b_k]$ contains a point x^* which minimizes $f_1 + f_2$.
- (ii) The derivative of f_1 at any minimizer of $f_1 + f_2$ and the derivative of f_1 and of $-f_2$ at $(a_k + b_k)/2$ belong to the interval $[c_k, d_k]$. (Note that the derivative of each f_i has to be constant on the set of minimizers of $f_1 + f_2$.)

At the first stage of the algorithm we start with $a_1 = 0$, $b_1 = 1$, $c_1 = -1$, and $d_1 = 1$. At the k th stage, the processors do the following: processor P_1 transmits a message $m_{1,k} = 0$ if $(-1)^{k-1}f'_1((a_k + b_k)/2) \leq (c_k + d_k)/2$; otherwise it transmits $m_{1,k} = 1$.

If $m_{1,k} = 0$ and $m_{2,k} = 1$, then $f'_1((a_k + b_k)/2) + f'_2((a_k + b_k)/2) \leq 0$. We may then let $a_{k+1} = (a_k + b_k)/2$ and leave b_k, c_k, d_k unchanged. Similarly, if $m_{1,k} = 1$ and $m_{2,k} = 0$, we let $b_{k+1} = (a_k + b_k)/2$ and leave a_k, c_k, d_k unchanged.

We now consider the case $m_{1,k} = m_{2,k} = 1$. Let x^* be a minimizer of $f_1 + f_2$ belonging to $[a_k, b_k]$. If $x^* \geq (a_k + b_k)/2$, then $f'_1(x^*) \geq f'_1((a_k + b_k)/2) \geq (c_k + d_k)/2$. If $x^* \leq (a_k + b_k)/2$, then $f'_2(x^*) \leq -f'_2((a_k + b_k)/2) \leq (c_k + d_k)/2$. In either case, we may let $c_{k+1} = (c_k + d_k)/2$ and leave a_k, b_k, d_k unchanged. Finally, if $m_{1,k} = m_{2,k} = 0$, a similar argument shows that we may let $d_{k+1} = (c_k + d_k)/2$ and leave a_k, b_k, c_k unchanged.

For each of the four cases, we see that a_k, \dots, d_k will preserve properties (i), (ii), which were postulated earlier. Furthermore, at each stage, either $b_k - a_k$ or $d_k - c_k$ is halved. Therefore, after at most $k = 2 \log(1/\epsilon)$ stages, we reach a point where either $b_k - a_k \leq \epsilon$ or $d_k - c_k \leq \epsilon$. If $b_k - a_k \leq \epsilon$, then there exists a minimizer which is within ϵ of a_k ; given that the derivative of $f_1 + f_2$ is bounded by one, it follows that $f_1(a_k) + f_2(a_k)$ comes within ϵ of the optimum, as desired. Alternatively, if $d_k - c_k \leq \epsilon$, then $|f'_1((a_k + b_k)/2) + f'_2((a_k + b_k)/2)| \leq d_k - c_k \leq \epsilon$. It follows that for any $\lambda \in [0, 1]$, we have $f_1(\lambda) + f_2(\lambda) \geq f_1((a_k + b_k)/2) + f_2((a_k + b_k)/2) - \lambda - (a_k + b_k)/2 \geq \epsilon$, which shows that $(f_1 + f_2)((a_k + b_k)/2)$ comes within ϵ of the optimum. ■

V. AN ALMOST OPTIMAL PROTOCOL FOR STRONGLY CONVEX PROBLEMS

We consider here the class $\tilde{\mathcal{F}}_{\infty, n, L}$ of strongly convex functions which was defined in Section III as the set of continuously differentiable convex functions satisfying (2.1)–(2.2). In this section we show that a suitable distributed version of the gradient projection algorithm comes close to the lower bound of Proposition 2.3, within an $O(\log n)$ factor. In particular,

for any fixed dimension n , we have a protocol whose dependence on ϵ is optimal.

In the protocol to be considered each processor computes the same sequence $\{x_k\}$ of elements of $[0, 1]^n$ according to the iteration

$$x_{k+1} = [x_k + \gamma g_k], \quad x_0 = 0. \quad (5.1)$$

We use the notation $[x]$ to denote the projection (with respect to the Euclidean metric) of a vector $x \in \mathbb{R}^n$ onto the convex set $[0, 1]^n$. Also, γ is a positive scalar stepsize and g_k is an approximation of the gradient of $f_1 + f_2$, evaluated at x_k . In particular, we let $g_k = f_1'(x_k) + f_2'(x_k)$ and we require that x_k satisfy

$$\|x_k - g_k\| \leq n^{1/2} \alpha^k, \quad (5.2)$$

where α is some positive constant, independent of k , belonging to $(0, 1)$. Naturally, we will have to ensure that there is enough communication so that each processor knows x_k at the beginning of the k th stage.

We start by estimating the number of steps required by the above algorithm to come to a small neighborhood of the optimal point. The argument is very similar to the standard proof that the gradient projection algorithm has a linear rate of convergence (Nemirovsky and Yudin, 1983, pp. 258–260) except that we need to take care of the fact that we use x_k instead of the exact gradient g_k . We denote by x^* the unique vector in $[0, 1]^n$ which minimizes $f_1 + f_2$ over that domain. (Uniqueness is a consequence of strict convexity, which follows from strong convexity.)

PROPOSITION 5.1. *If $f \in \mathcal{F}_{SC, ML}$, if x_k, x_* satisfy (5.1)–(5.2), if the stepsize γ is small enough, and if α is sufficiently close to 1, then there exist $A, B, C > 0$, depending only on M, L , such that:*

$$(i) \quad f(x_k) - f(x^*) \leq A\alpha^n, \quad (5.3)$$

$$(ii) \quad \|x_k - x^*\|^2 \leq Bn\alpha^{2n}, \quad (5.4)$$

$$(iii) \quad \|x_k - x_*\| \leq Cn^{1/2}\alpha^n. \quad (5.5)$$

Proof. We will prove the result with the following choices of constants: we let $\gamma = 1/(ML)$, $B = 2A/L$, and $C = 2B^{1/2}$. The constants A and α will be fixed later.

We state without proof the following properties of functions in $\mathcal{F}_{SC, ML}$ (Nemirovsky and Yudin, 1983, pp. 254–255):

$$(i) \quad f(x) - f(y) \leq ML\|x - y\|, \quad (5.6)$$

$$(ii) f(x + y) \geq f(x) + f'(x)y + (L/2)|y|^2, \quad (5.7)$$

$$(iii) f(x + y) \leq f(x) + f'(x)y + (LM/2)|y|^2. \quad (5.8)$$

We will be also using the inequality

$$f'(x^*)y - x^*y \geq 0, \quad \forall y \in [0, 1]^n, \quad (5.9)$$

which is a necessary and sufficient condition for optimality of x^* .

We continue with the main part of the proof, which proceeds by induction on k . We first show that part (i) holds for $k = 0$. Using the convexity of f , we have

$$f(x^*) \leq f(x_0) + f'(x_0)(x^* - x_0) \leq f(x_0) + f'(x_0)^T(x^* - x_0).$$

Using (2.2), we see that $f'(x_0)$ is bounded by $MLn^{1/2}$; also, $\|x^* - x_0\| = \|x^*\|$ is bounded by $n^{1/2}$. It follows that $f(x_0) - f(x^*) \leq MLn \leq An$, as long as A is chosen larger than ML .

Suppose now that (5.3) is valid for some nonnegative integer k . Using (5.7) and then (5.9) we obtain

$$\begin{aligned} f(x_0) &\geq f(x_0) + f'(x^*)(x_0 - x^*) + \frac{L}{2}\|x_0 - x^*\|^2 \\ &\geq f(x^*) + \frac{L}{2}\|x_0 - x^*\|^2. \end{aligned} \quad (5.10)$$

We now use (5.10) and the induction hypothesis to obtain

$$\|x_0 - x^*\|^2 \leq \frac{2}{L}[f(x_0) - f(x^*)] \leq \frac{2}{L}Ana^{1/2} = Bna^{1/2}. \quad (5.11)$$

We have therefore shown that (5.4) is also valid for that particular k . We then use (5.4) and the triangle inequality to obtain

$$\|x_{k+1} - x^*\| \leq \|x_{k+1} - x_k\| + \|x_k - x^*\| \leq 2B^{1/2}n^{-1/2}a^{1/2}, \quad (5.12)$$

which proves (5.5) for that same value of k .

We now prove (5.3) for $k + 1$, which will complete the induction. Using the definition of the projection, x_{k+1} minimizes $\|y - x_k\| + \gamma\|y\|^2$ over $y \in [0, 1]^n$, which is equivalent to minimizing

$$f(x_k) + (x_k, y - x_k) + \frac{1}{2\gamma}\|y - x_k\|^2 \quad (5.13)$$

over all $x \in [0, 1]^n$. Let us use the notation $J_k(x)$ to denote the expression (5.13) as a function of x . Let $z = x_k + (1/M)(x^* - x_k)$. Note that $z \in [0, 1]^n$ because x_k, x^* belong to $[0, 1]^n$. Thus, by the minimizing property of x_k , we have

$$J_k(x_{k+1}) \leq J_k(z). \quad (5.14)$$

Now,

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + (g_k, x_{k+1} - x_k) + \frac{LM}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) + (g_k, x_{k+1} - x_k) + \frac{LM}{2} \|x_{k+1} - x_k\|^2 + (x_k - g_k, x_{k+1} - x_k) \\ &\leq J_k(x_{k+1}) + (n^{1/2}\alpha^4)(2B^{1/2}n^{1/2}\alpha^4) \\ &\leq J_k(z) + 2B^{1/2}n\alpha^2 \\ &= f(x_k) + (g_k, \frac{1}{M}(x^* - x_k)) + \frac{LM}{2} \frac{1}{M^2} \|x^* - x_k\|^2 + 2B^{1/2}n\alpha^2 \\ &\leq f(x_k) + (g_k, \frac{1}{M}(x^* - x_k)) + \frac{L}{2M} \|x^* - x_k\|^2 + 2B^{1/2}n\alpha^2 \\ &\quad + (x_k - g_k, \frac{1}{M}(x^* - x_k)) \\ &\leq (1 - \frac{1}{M}) f(x_k) + \frac{1}{M} \left[f(x_k) + (g_k, x^* - x_k) + \frac{L}{2} \|x^* - x_k\|^2 \right] \\ &\quad + 3B^{1/2}n\alpha^2 \\ &\leq (1 - \frac{1}{M}) f(x_k) + \frac{1}{M} f(x^*) + 3B^{1/2}n\alpha^2. \end{aligned}$$

Here, the first inequality followed from (5.8); the second from the Schwarz inequality; the third from (5.2), (5.12), and the definition of $J_k(x_{k+1})$; the fourth from (5.14). In the equality we made use of the definition of z and J_k , and the next step followed from the Schwarz inequality; then, we used the fact $M \geq 1$, (5.2), and (5.11); finally, the last line followed from (5.7). We therefore have, using the induction hypothesis,

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq (1 - \frac{1}{M}) (f(x_k) - f(x^*)) + 3B^{1/2}n\alpha^2 \quad (5.15) \\ &\leq (1 - \frac{1}{M}) A n \alpha^2 + 3 \left(\frac{2A}{L} \right)^{1/2} n \alpha^2. \end{aligned}$$

The induction will be completed if the right-hand side of (5.15) is smaller than $An\alpha^{2^{k+1}+1}$. This is accomplished by taking $\alpha \in (0, 1)$ close enough to 1 so that $1 - 1/M < \alpha^2$ and then choosing A large enough so that the term involving $A^{1/2}$ is negligible in comparison with the first term in the right-hand side of (5.15). This concludes the proof. ■

We now return to the distributed protocol. Since $t_1, t_2 \in \tilde{C}(\tilde{S}_0, \alpha, \epsilon)$, it follows that $t_1 + t_2 \in \tilde{C}(\tilde{S}_0, \alpha, 2\epsilon)$. Consequently, Proposition 5.1 applies to $t_1 + t_2$ and shows that after $O(\log(1/\epsilon) + \log n)$ stages, the algorithm (5.1)–(5.2) reaches a point which is within ϵ from optimality.

We now indicate how the protocol may be implemented with $O(n \log n)$ bits being communicated at each stage. All we need to do is to make sure that the processors share enough information at each stage to be able to compute a vector x_k satisfying (5.2). This is accomplished by letting each processor know a set of scalars $x_i(u, j)$, $i = 1, 2$, $j = 1, \dots, n$, such that $x_i(u, j) - g_i(u, j) \leq \alpha^i$, where $g_i(u, j)$ is the j th component of $f_i(x_i)$. We first consider stage $k = 0$. Using (2.2) we see that $g_i(u, j)$ is bounded by $O(n^{1/2})$, for each i, j . Therefore, it is sufficient to transmit $O(\log n)$ bits, to specify each component with accuracy $\alpha^0 = 1$.

Suppose now that $k > 0$ and that quantities $x_i(u, j)$ with the desired properties have been shared at stage $k - 1$. We have $g_i(u, j) - x_i(u, j) + g_i(u, j) - x_i(u, j) + g_i(u, j) - x_i(u, j) + x_i(u, j) - x_i(u, j) \leq LM|x_i - x_i|^2 + n^2\alpha^{2(k-1)} \leq (LMCn^{1/2} + n^2)\alpha^{2(k-1)}$. (Here we have made use of (5.6), our hypothesis that x_i satisfies (5.2), and part (ii) of Proposition 5.1.) Let us impose the additional requirement that $x_i(u, j)$ be an integer multiple of α^k . This requirement does not prohibit the attainment of our goal, which is to satisfy inequality (5.2). With this requirement, there are at most $\alpha^{-k}(LMCn^{1/2} + 1) + 1$ possible choices for $x_i(u, j)$. Therefore, each processor P_i may choose $x_i(u, j)$ as above and transmit its value to the other processor, while communicating only $O(\log n)$ bits for each component j , thus leading to a total of $O(n \log n)$ communications per stage. We have thus proved the following result.

PROPOSITION 5.2. For any fixed M, L , we have $C(\tilde{C}(\tilde{S}_0, \alpha, \epsilon), \epsilon) \leq O(n \log n \log n + \log(1/\epsilon))$.

VI. POSSIBLE EXTENSIONS AND OPEN QUESTIONS

1. The protocol of Section V is likely to be far from optimal concerning the dependence on the parameters M and L . The gradient algorithm tends to be inefficient for poorly conditioned problems (large M), as opposed to variations of the conjugate gradient method (Nemirovsky and Yudin, 1983). It remains to be seen whether a suitable approximate version of the conjugate gradient method admits a distributed implementation with low communication requirements as a function of M .

2. For the class \mathcal{F}_1 , gradient methods do not work and the gap between the lower bound of Section II and the upper bound of Section III remains open. We believe that the factor of n^2 in the upper bound cannot be reduced. The reason is that any conceivable algorithm would need to consider at least $O(n \log(1/\epsilon))$ points and it is hard to imagine any useful transfer of information concerning the behavior of the function in the vicinity of a point which does not require $O(n)$ messages. On the other hand, it may be possible to reduce the factor $\log(1/\epsilon)$ to just $\log(1/\epsilon)$ although we do not know how to accomplish this. A related open problem concerns the $O(\log n)$ gap between Propositions 5.2 and 2.3, for the class $\mathcal{F}_{\text{SCVLC}}$.

3. Some directions along which it is likely that the results can be extended concern the case of $K \geq 2$ processors and the case where the constraints under which the optimization is carried out are not commonly known; for example, we may have a constraint of the form $g_1(x) + g_2(x) \leq 0$, where each g_i is a convex function known by processor P_i .

REFERENCES

- ARLSON, H. (1980), Lower bounds on information transfer in distributed computations, *J. Assoc. Comput. Mach.*, **27**, No. 2, 384-392.
- ARIE, A. V., ULLMAN, J. D. and YANNAKAKIS, M. (1983), On notions of information transfer in VLSI circuits, in: *Proceedings, 15th STOC*, 1983, pp. 133-139.
- AWERBUCH, B. and GALLAGER, R. G. (1985), "Communication Complexity of Distributed Shortest Path Algorithms", Technical Report LIDS-P-1473, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.
- MEHLHORN, K. and SCHMIDT, E. M. (1982), Las Vegas is better than Determinism in VLSI and distributed computing, in: *Proceedings, 14th STOC*, 1982, pp. 331-337.
- NEMIROVSKY, A. S. and YUDIN, D. B. (1983), "Problem Complexity and Method Efficiency in Optimization," Wiley, New York.
- PANG, K. F. and EL GAMAL, A. (1986), Communication complexity of computing the Hamming distance, *SIAM J. Comput.*, **15**, No. 4, 932-947.
- PAPADIMITROU, C. H. and SPIER, M. (1982), Communication complexity, in: *Proceedings, 14th STOC*, 1982, pp. 196-200.
- PAPADIMITROU, C. H. and STEIGLITZ, K. (1982), *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ.
- PAPADIMITROU, C. H. and TSITSIKIS, J. N. (1982), On the complexity of designing distributed protocols, *Inform. and Control*, **53**, No. 3, 211-218.
- TRAUB, J. F. and WOZNIAKOWSKI, H. (1980), "A General Theory of Optimal Algorithms," Academic Press, Orlando, FL.
- ULLMAN, J. D. (1984), "Computational Aspects of VLSI," Comput. Sci. Press, Rockville, MD.
- YAO, A. C. (1979), Some complexity questions related to distributed computing, in: *Proceedings, 11th STOC*, 1979, pp. 209-213.

END

DATE

FILMED

6-88

DTIC